Data Sources and Analysis for "The Pandemic Tsunami: How COVID-19 Swept Across America"

Jason M. Barr, Rutgers University-Newark
Troy Tassier, Fordham University

February 24, 2021

1. The Regression Model

The dependent variable for all regressions is the (natural) log 1 plus the total number of cases in county *i* at time *t*. The first set of models estimates the following equation,

$$\ln(cases_{it}) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots \beta_n x_{ni} + \varepsilon_{it},$$

where x_{ji} is variable *j=1,…,n,* for county *i*. Note all x's are except avg household size is in ln form and in cases when a variable had zeros we took ln(1+x). These control variables, at the county level, and their sources, are as follows:

- *Population*: American Community Survey, 2018. Table PEP_2018_PEPANNRES_with_ann.
- *Land area*: From U.S. County Shapefile from the Census Bureau: https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html
- - Density is population divided land area.
- *Gross Domestic Product*: 2018, https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas
- *Number of days since first case*: (same as Coronavirus cases source below—USAFacts.org).
- *Race*: %Black, %Asian, %Hispanic, % Native American % Hawaiian or Pacific Islander, %Other Race, and %Two or More Races (note that %White is the omitted variable): American Community Survey 2018, 5-year average. Table ACSDT5Y2018.B02001_data_with_ov.
- *% below poverty line*. American Community Survey, 2018, 5-year average. Table ACSST5Y2018.S1701_data_with_ove.
- *Avg*. House Hold Size. American Community Survey, 2019, Table ACSST5Y2019.S1101_data_with_ove

- Total reported COVID-19 cases at county level by each are from: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/

2. Regression Results

Note regressions were run for each month, where dependent variable is number of total cases up to the first day of that month, with the total number of cases one month prior as a control variable. Thus the coefficients on the other variables give the effect of those variables on the growth rate of COVID-19 cases over that month. Table 1 gives regressions from March 1, 2020 to August 1, 2020, respectively. Note average household size is in levels, all other variables in logs.

The figures in the post are the plots of the respective coefficient values for each month.

	(1) 3/1/20	(2) 4/1/20	(3) 5/1/20	(4) 5/1/20	(5) 7/1/20	(6) 8/1/20
avghhsize	-0.000244 (-0.03)	0.0946 (1.25)	0.325 (4.03)***	0.290 (5.25)***	0.202 (4.70)***	0.0613 (1.57)
lnDensity	-0.00717 (-3.01)***	0.387 (20.38)***	0.251 (11.36)***	0.124 (8.12)***	0.173 (14.53)***	0.204 (18.01)***
lnBlack	-0.00199 (-0.71)	0.140 (6.24)***	0.262 (10.90)***	0.116 (6.88)***	0.0241 (1.82)*	0.00897 (0.75)
lnNatAm	-0.00624 (-1.40)	0.0785 (2.21)**	0.0510 (1.34)	0.00918 (0.35)	0.0211 (1.05)	0.0565 (3.10)***
lnAsian	0.0242 (4.26)***	0.401 (8.88)***	0.0641 (1.31)	-0.0125 (-0.38)	0.0600 (2.35)**	-0.0221 (-0.95)
lnHIPI	0.0552 (2.84)***	-0.304 (-1.97)**	-0.344 (-2.09)**	0.228 (2.02)**	0.319 (3.66)***	0.197 (2.49)**
lnOtherRace	-0.00892 (-0.54)	0.378 (2.87)***	-0.156 (-1.11)	-0.163 (-1.70)*	-0.148 (-1.99)**	0.0307 (0.46)
lnTwoorMor~s	-0.00469 (-0.68)	-0.0655 (-1.20)	-0.0723 (-1.24)	-0.0649 (-1.63)	-0.0924 (-3.00)***	-0.0408 (-1.45)
lnHispanic	-0.00879 (-2.37)**	0.110 (3.72)***	0.246 (7.81)***	0.237 (10.85)***	0.160 (9.24)***	0.0839 (5.25)***
lnPoverty%	0.00831 (1.33)	-0.259 (-5.23)***	-0.0343 (-0.64)	0.0830 (2.29)**	0.139 (4.95)***	0.0750 (2.95)***
tfirstcase	0.00317 (11.91)***	0.0731 (33.55)***	0.000451 (0.16)	-0.00371 (-2.24)**	0.00300 (2.41)**	0.00481 (4.27)***
Lncovid_3/1		-0.480 (-2.82)***				
Lncovid_4/1			0.691 (30.16)***			
Lncovid_5/1				0.846 (68.61)***		
Lncovid_6/1					0.797 (85.47)***	
Lncovid_7/1						0.749 (80.32)***
_cons	-0.265 (-3.67)***	9.113 (15.83)***	6.802 (10.49)***	2.873 (6.39)***	4.462 (12.78)***	6.345 (19.34)***
N	2228	2228	2228	2228	2228	2228
R-sq	0.192	0.823	0.844	0.929	0.956	0.960
adj. R-sq	0.169	0.818	0.840	0.927	0.955	0.959
AIC	-4511.2	4720.5	5013.5	3319.6	2170.7	1740.4
BIC	-4442.7	4794.7	5087.8	3393.8	2244.9	1814.7

Dep. Var.: ln(1+Covid cases for county 1 on day one of each month). State-level dummy variables included in each regression. t statistics in parentheses * p<0.10, ** p<0.05, *** p<0.01.

.

	(1) 9/1/20	(2) 10/1/20	(3) 11/1/20	(4) 12/1/20	(5) 1/1/21	(6) 2/1/21
avghhsize	0.0487 (1.72)*	0.0610 (2.69)***	0.0104 (0.48)	-0.0227 (-1.07)	-0.0441 (-2.87)***	-0.00377 (-0.38)
lnDensity	0.103 (11.86)***	0.0383 (5.29)***	0.0536 (7.75)***	0.0752 (10.93)***	0.0354 (6.91)***	0.0141 (4.19)***
lnBlack	0.00138 (0.16)	-0.0259 (-3.79)***	-0.0316 (-4.90)***	-0.0329 (-5.22)***	-0.0293 (-6.46)***	-0.000837 (-0.29)
lnNatAm	0.0119 (0.91)	-0.0246 (-2.33)**	0.0111 (1.10)	0.0197 (1.99)**	0.000677 (0.09)	0.000320 (0.07)
lnAsian	-0.0201 (-1.20)	0.0440 (3.29)***	-0.0107 (-0.84)	-0.0113 (-0.90)	-0.0120 (-1.32)	0.00530 (0.90)
lnHIPI	0.0874 (1.53)	-0.0192 (-0.42)	0.00138 (0.03)	-0.0889 (-2.07)**	-0.00181 (-0.06)	0.0298 (1.47)
lnOtherRace	0.0128 (0.26)	0.0130 (0.33)	0.0521 (1.40)	-0.0213 (-0.58)	0.0356 (1.34)	-0.0136 (-0.79)
lnTwoorMor~s	-0.0386 (-1.91)*	0.0292 (1.80)*	0.0301 (1.95)*	0.0149 (0.98)	0.0432 (3.92)***	0.0165 (2.31)**
lnHispanic	-0.0304 (-2.63)***	-0.0541 (-5.91)***	-0.0203 (-2.36)**	-0.00631 (-0.75)	-0.0287 (-4.73)***	-0.0172 (-4.39)***
lnPoverty_~t	0.0583 (3.17)***	0.0615 (4.16)***	0.0452 (3.21)***	-0.00867 (-0.62)	-0.0313 (-3.11)***	-0.0338 (-5.18)***
tfirstcase	0.000775 (0.95)	0.000131 (0.20)	0.000275 (0.44)	0.000601 (0.99)	0.000989 (2.25)**	0.00117 (4.10)***
Lncovid_8/1	0.872 (112.26)***					
Lncovid_9/1		0.929 (140.45)***				
Lncovid_10/1			0.910 (141.41)***			
Lncovid_11/1				0.887 (134.46)***		
Lncovid_12/1					0.944 (185.03)***	
Lncovid_1/1						0.974 (286.12)***
_cons	3.495 (13.75)***	1.480 (6.96)***	2.191 (10.70)***	3.248 (15.75)***	1.906 (12.13)***	0.879 (8.36)***
N	2228	2228	2228	2228	2228	2228
R-sq	0.977	0.983	0.982	0.980	0.989	0.996
adj. R-sq	0.976	0.983	0.982	0.980	0.989	0.995
AIC	284.6	-688.0	-904.1	-982.8	-2411.3	-4334.4
BIC	358.9	-613.8	-829.8	-908.5	-2337.1	-4260.2

Dep. Var.: ln(1+Covid cases for county 1 on day one of each month). State-level dummy variables included in each regression. t statistics in parentheses * $p<0.10$, ** $p<0.05$, *** $p<0.01$.